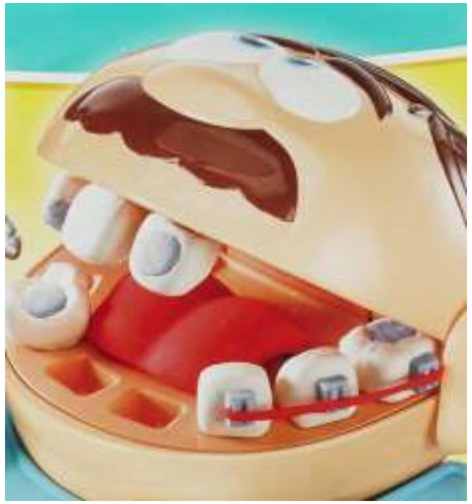
Uncertainty

ITI0210, lecture 9 (2021)

Logic and Reality

Logic is rigid:

$$Toothache \rightarrow Cavity$$



In logic, this is **always** true

In real world, it is true **sometimes**, **frequently**, etc

Fudge Factors

First intuition: add some confidence value

```
toothache(X) -> cavity(X) : 0.8
```

I'm 80% confident that this diagnosis is true;

can come from e.g. medical statistics

Fudge Factors

Not as easy as it looks

```
smokes(X) & roommate(X,Y) -> smokes(Y) : 0.3
smokes(X) & influences(X,Y) -> smokes(Y) : 0.7
smokes(bob).
roommate(bob, carl).
influences(bob, carl) : 0.5
```

It's not obvious how to compute confidence for `smokes(carl)`.

For example: does the rule with 0.3 confidence increase or decrease total confidence?

Conditional Probability

Probability has a **defined meaning** and lots of **established theory**, **software** etc

$$P(cavity|toothache) = 0.8$$

Means, “given the evidence: toothache, probility of cavity is 0.8”

Does **not** mean “if toothache then 0.8 chance of cavity”.

$$P(cavity|toothache, cavity) = 1$$

Probability can change, given different evidence

Probability

The Basics

Axioms

Probability space Ω, atomic events $\omega \in \Omega$

1. $P(\omega) \geq 0$
2. $\sum_{\omega \in \Omega} P(\omega) = 1$

Let's say an "event" A is a set of atomic events

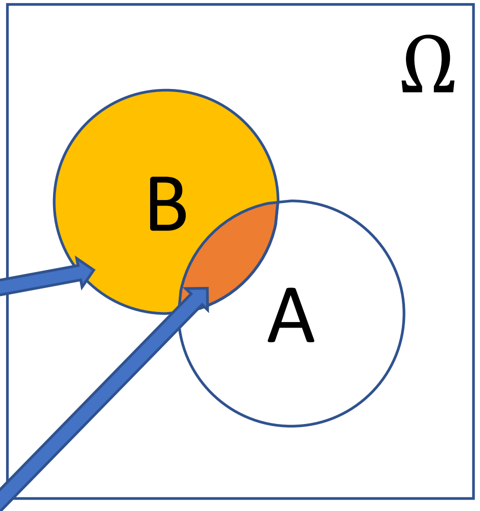
Example: one heads, one tail

3. $P(A) = \sum_{\omega \in A} P(\omega)$

Definitions

Conditional probability

$$P(A|B) = \frac{P(A,B)}{P(B)}$$



B is known or given

Reading $P(A, B)$: probability that A and B are both true

(Imagine that area expresses probability)

Together with the axioms, this is all the theory we need!

Random variable

Examples:

r.v. Odd says the roll of a six-sided die was odd.

$$P(Odd) = \sum_{\omega \in Odd} P(\omega) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 0.5$$

Can be numeric (e.g. $Temperature$), categorical, etc:

$$Weather \in \{sunny, cloudy, rain, snow\}$$

Random variable

Training data for machine learning can be considered random variables

X_1	X_2	X_3	X_4	Y
sepal l	sepal w	petal l	petal w	class
4.6	3.2	1.4	0.2	*Iris setosa*
5.3	3.7	1.5	0.2	*Iris setosa*
5.0	3.3	1.4	0.2	*Iris setosa*
7.0	3.2	4.7	1.4	*Iris versicolor*
6.4	3.2	4.5	1.5	*Iris versicolor*

The "Iris Flower Dataset"

Samples

Sampling random variables:

- throw a coin, see which side is up
- pick an iris flower, measure its petals

Samples can be biased

But “bent coin” may not be the best example:

https://izbicki.me/blog/how-to-create-an-unfair-coin-and-prove-it-with-math.html

Book conventions

$P(Weather = sunny) = 0.3$ probability of a given value for variable

$P(cavity) = 0.3$ probability that the Boolean variable $Cavity$ is true

$P(\neg cavity) = 0.3$ probability that $Cavity = false$

$P\big(a \vee (b \wedge c)\big) = 0.01$ probability for a **proposition** being true

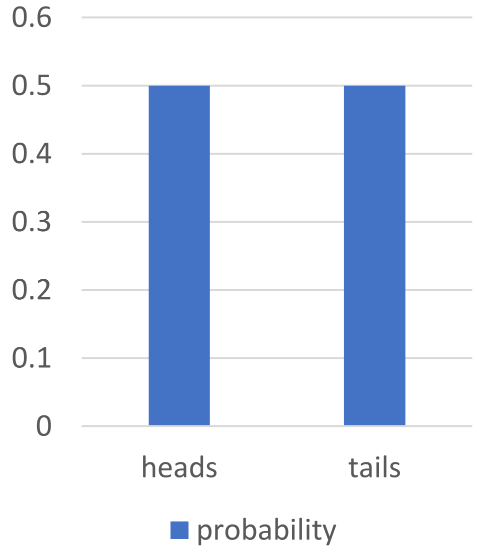
a, b is the same as $a \wedge b$

Distributions

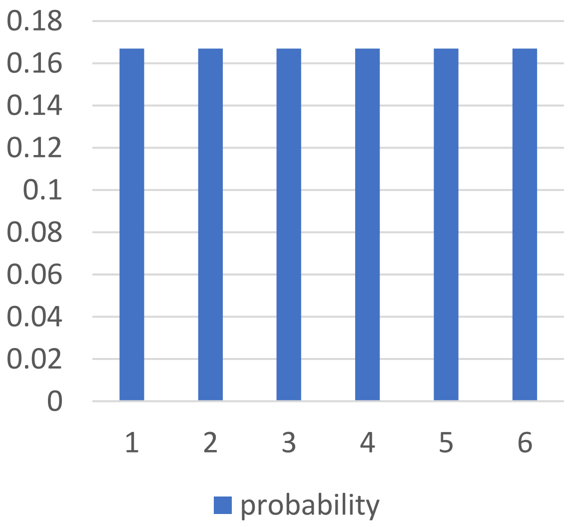
Where do probabilities come from

Examples

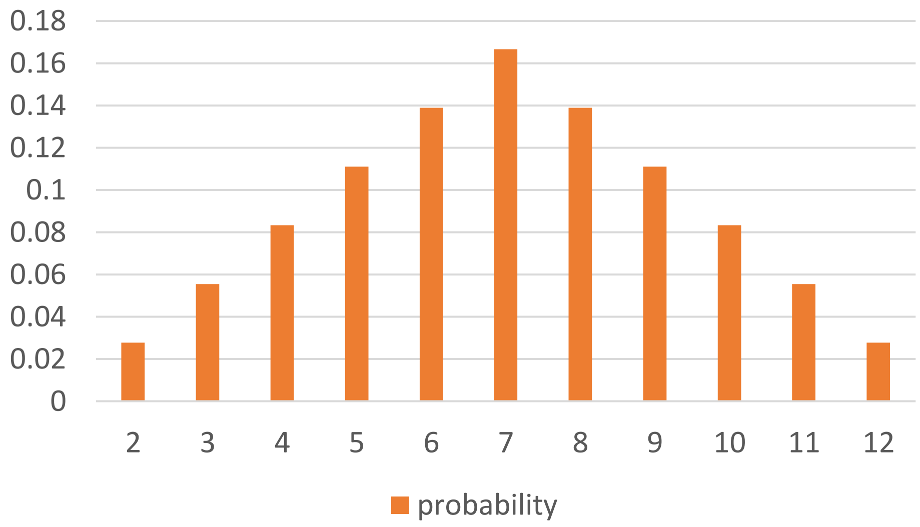
Theoretical coin



6-sided die



Sum of two dice

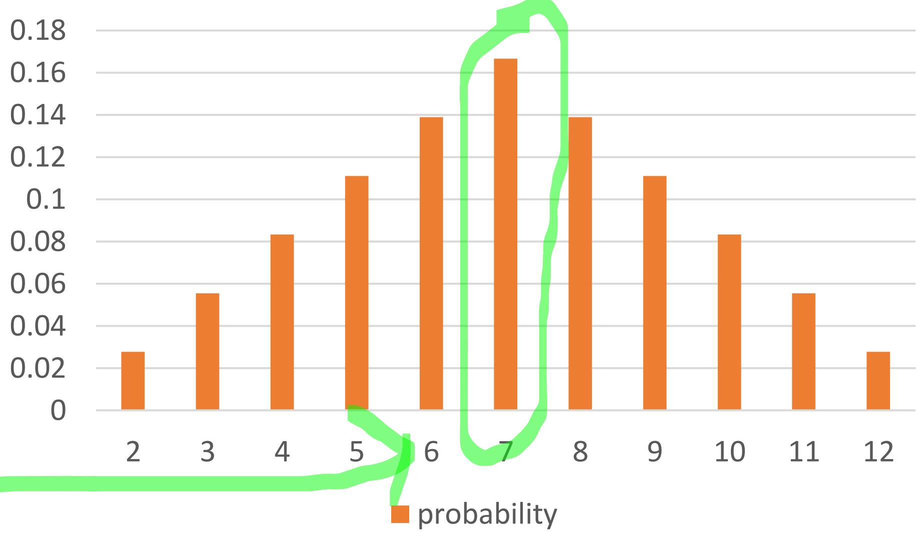


Using a distribution

The distribution tells us

what is $P(X = x)$ for some random variable

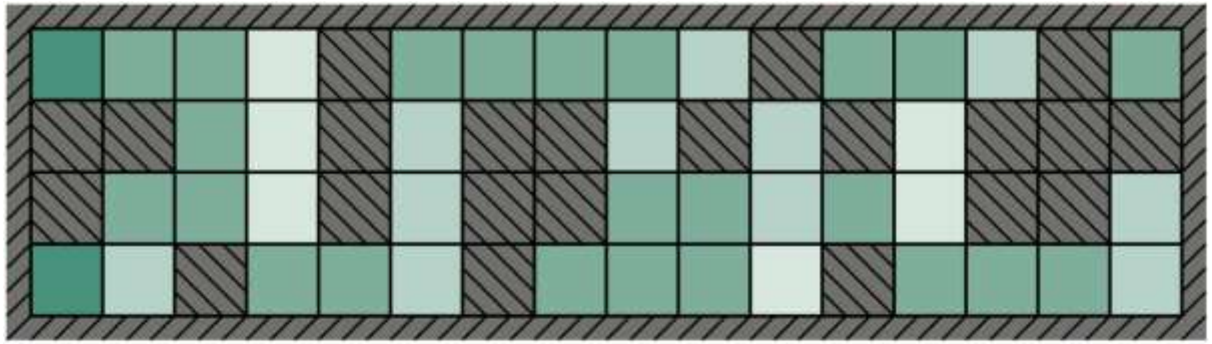
Example: what is the probability of throwing 2 dice and getting a "7"?



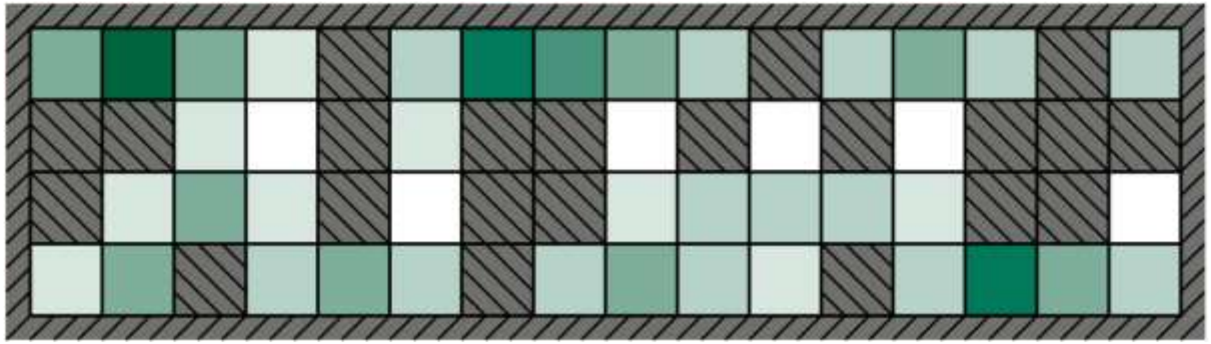
Answer: 0.167

Using a distribution

Robot observes walls N, S, W. Distribution of estimated robot location:

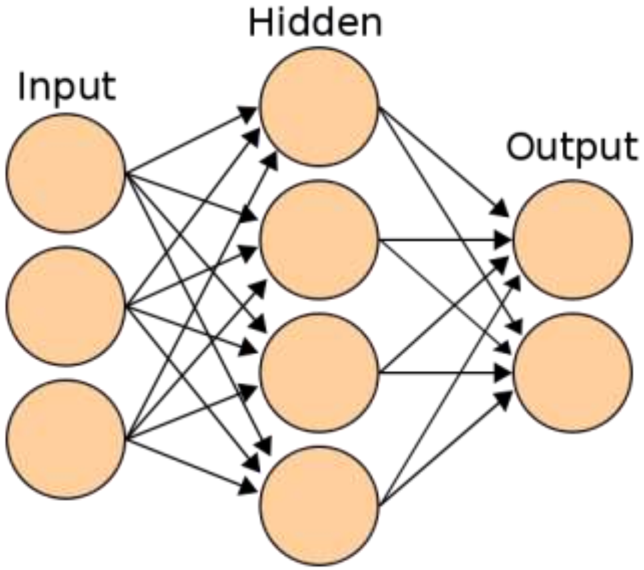


Move east, observe walls N, S:



Using a distribution

The “softmax” output of a neural network
is a probability distribution



x_i	softmax $\frac{e^{x_i}}{\sum_j e^{x_j}}$	class
-0.12	0.107	cat
2	0.893	dog
SUM:	1.0	

Distributions and Models

Let’s say you tossed a coin 100 times

result 52 heads/48 tails

Empirical distribution $P(heads) = 0.52$

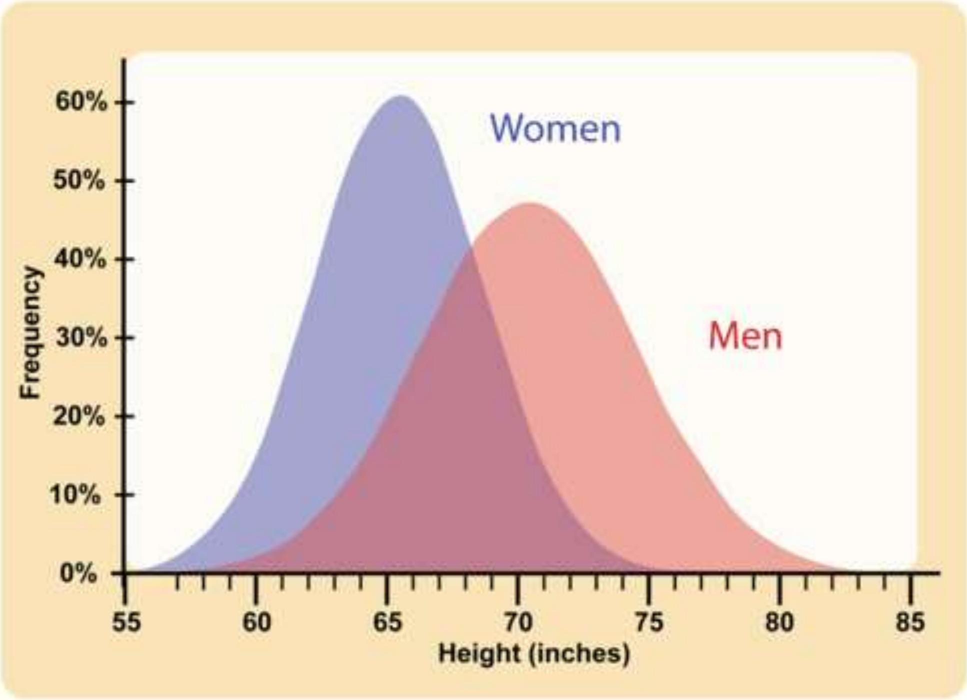
If you continue long enough, $P(heads)$ will approach 0.5, right?

Actually, $P(heads) = 0.5$ is an idealized model

It is not more “real” than your experiment

Distributions and Models

Gaussian height distribution: it is an **approximation** of reality



Images: ck12.org / Simon Fraser University

Full Joint Distribution

Gives the probability for **each combination** of **all** variables

It's not important **how** the probability is determined

e.g. could be simply a table:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Discrete variables

Sum is 1

Probabilistic Inference

Using the Full Joint Distribution

Marginal probability

The full joint distribution is incredibly useful for answering questions about probabilities

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

e.g. "What is the overall probability of toothaches?"

Using axiom 3

$$P(toothache) = \sum_{\omega \in T} P(\omega) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

T is the set of atomic events where $Toothache = true$

Conditioning

Or with conditional probability:

"Known symptoms: toothache.

What is the probability of **no cavity**?"

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(\neg cavity|toothache) = \frac{P(\neg cavity, toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Using

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Probabilistic Inference

Given evidence $\boldsymbol{e}$ and query variables $\boldsymbol{q}$

Note the bold, these are vectors, i.e. $\boldsymbol{e} = \{e_1, \ldots, e_k\}$

$$P(\boldsymbol{q}|\boldsymbol{e}) = \frac{P(\boldsymbol{q}, \boldsymbol{e})}{P(\boldsymbol{e})} = \frac{\sum_{\boldsymbol{h}'} P(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{h}')}{\sum_{\boldsymbol{q}', \boldsymbol{h}'} P(\boldsymbol{q}', \boldsymbol{e}, \boldsymbol{h}')}$$

$\boldsymbol{h}'$ are all possible values for the rest of the random variables

$\boldsymbol{q}'$ are all possible values for the query variables

If this formula confuses you, go back to previous slide (same formula).

Probabilistic Inference

We now have something similar to logical inference:

given evidence, we can ask questions

$$P\left(pit_{2,1} \middle| \underbrace{breeze_{2,2}, \neg breeze_{1,1}}_{\text{Facts}}\right)$$

Query

$$P(\underbrace{fraud}_{\text{Query}} | \underbrace{TotalEUR > 10000, CountryBlacklisted = true}_{\text{Facts}})$$

Bonus: supports **uncertainty**, unlike logical inference

Mini-Tutorial

How to compute formulas like this

$$P(\boldsymbol{q}|\boldsymbol{e}) = \frac{P(\boldsymbol{q},\boldsymbol{e})}{P(\boldsymbol{e})} = \frac{\sum_{\boldsymbol{h'}} P(\boldsymbol{q},\boldsymbol{e},\boldsymbol{h'})}{\sum_{\boldsymbol{q'},\boldsymbol{h'}} P(\boldsymbol{q'},\boldsymbol{e},\boldsymbol{h'})}$$

Test data: http://lambda.ee/w/images/7/74/Somerville.txt

(adapted from https://archive.ics.uci.edu/ml/datasets/Somerville+Happiness+Survey)

Python example code: coming